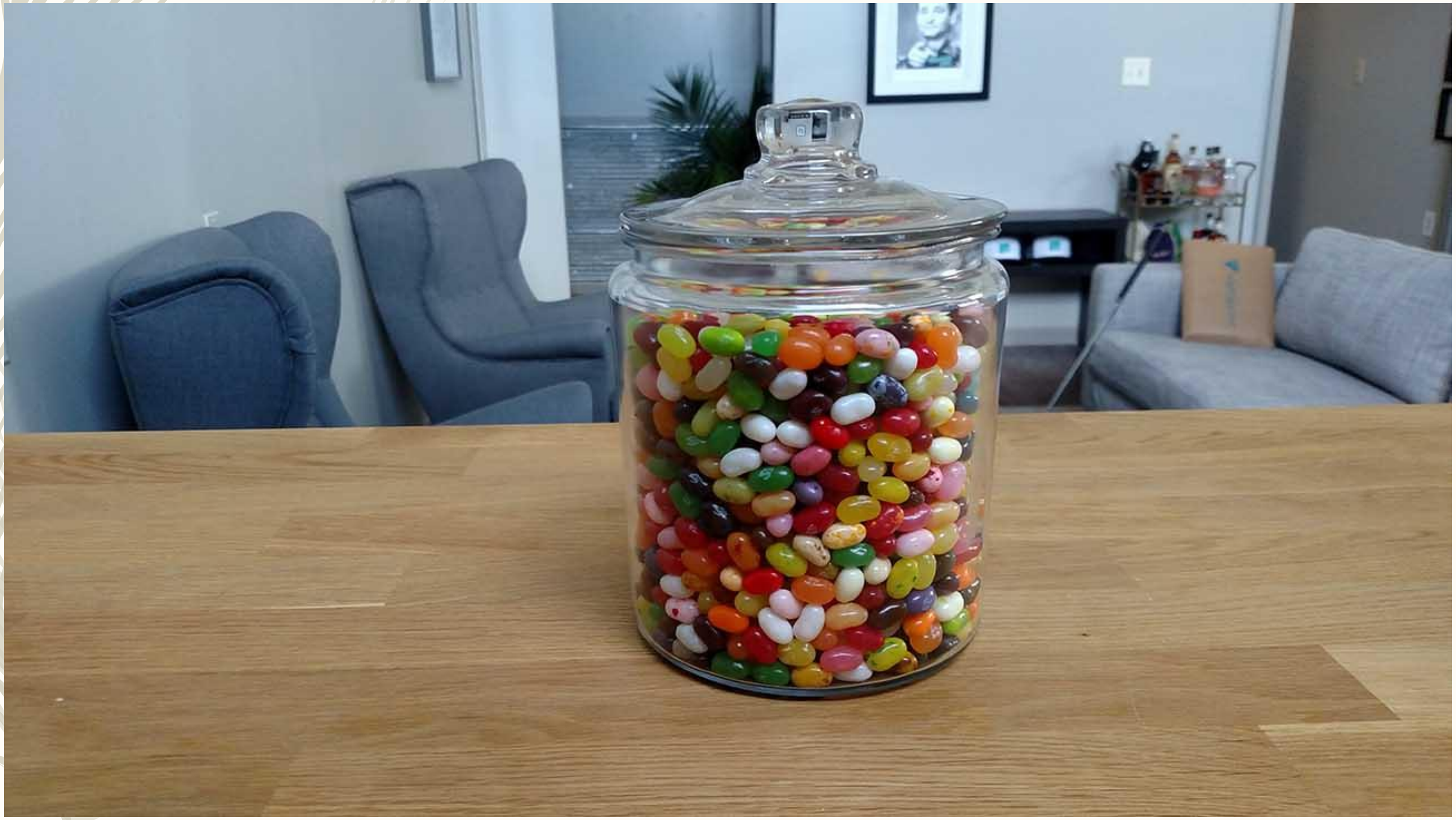
Goodies*

* Goodies related to animals, plants and numbers…

"… Estive a pensar e lembrei-me de lhe perguntar como é que se consegue estimar o número de feijões, se solicitarmos a um grande número de pessoas que façam a melhor estimativa sobre a quantidade de feijões existentes num frasco. Vi algures que a estimativa tende sempre para um número aproximado do valor exato e queria saber porquê…"

Aluno anónimo, e-mail ontem Mon 9/30/2019 4:15 PM

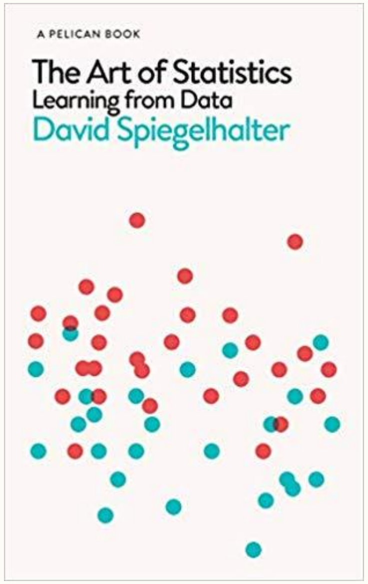


John Kay JULY 24 2012 64

In 1906, the great statistician Francis Galton observed a competition to guess the weight of an ox at a country fair. Eight hundred people entered. Galton, being the kind of man he was, ran statistical tests on the numbers. He discovered that the average guess (1,197lb) was extremely close to the actual weight (1,198lb) of the ox. This story was told by James Surowiecki, in his entertaining book *The Wisdom of Crowds*.

https://www.ft.com/content/bfb7e6b8-d57b-11e1-af40-00144feabdc0

Also a good description of this in the outstanding book:
(statistics for the wider public!)



Jelly bean challenge: a popular game at kids parties! Even seen on twitter… etc

"… Estive a pensar e lembrei-me de lhe perguntar como é que se consegue estimar o número de feijões, se solicitarmos a um grande número de pessoas que façam a melhor estimativa sobre a quantidade de feijões existentes num frasco. Vi algures que a estimativa tende sempre para um número aproximado do valor exato e queria saber porquê…"

Figura 1 – Jelly beans

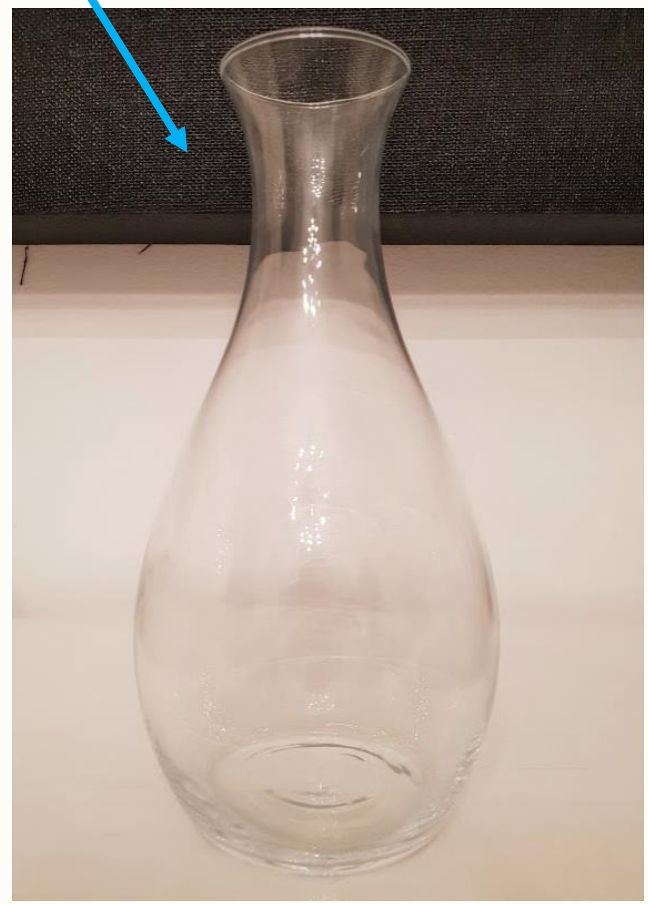
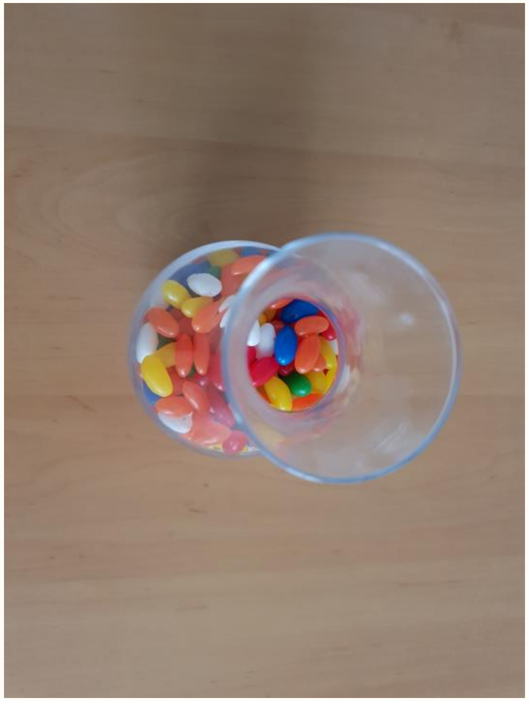


Figura 2 – frasco

Realizamos a nossa própria experiência, que de certa forma nos mostra o teorema do Limite Central em ação. Se não houver nenhum viés consistente, vamos conseguir estimar o número de *jelly beans* de forma razoável… parece magia…!

Sem falarem entre vocês, olhem para a jarra e estimem o número de jelly beans. Escrevam num pedaço de papel e venham colocar em cima da mesa. Eu contei e sei quantos são!



Realizamos a nossa própria experiência, que de certa forma nos mostra o teorema do Limite Central em ação. Se não houver nenhum viés consistente, vamos conseguir estimar o número de *jelly beans* de forma razoável… parece magia…!

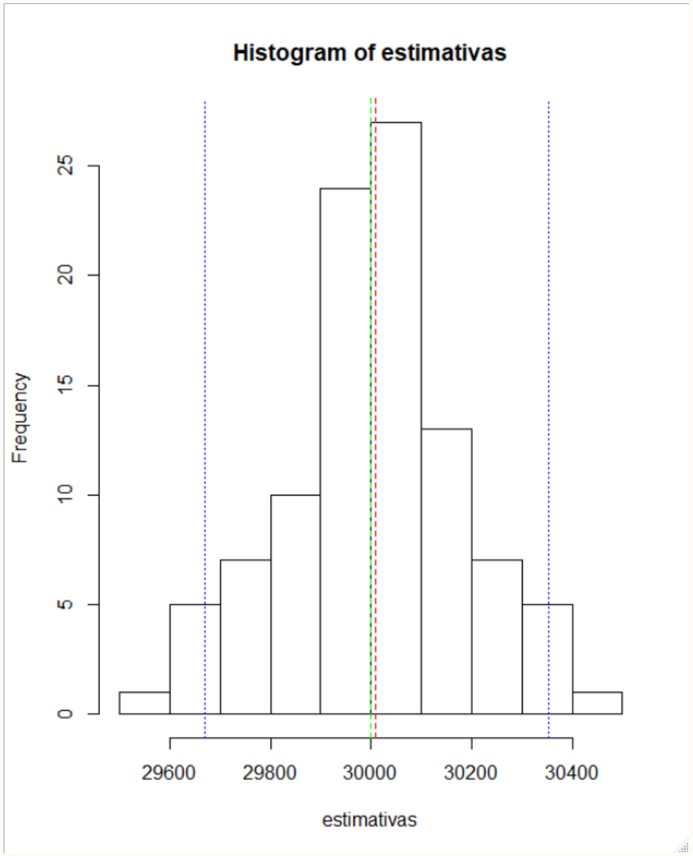
Sem falarem entre vocês, olhem para a jarra e estimem o número de jelly beans. Escrevam num pedaço de papel e venham colocar em cima da mesa. Eu contei e sei quantos são!

Como todo o bom mágico, hoje tenho um ajudante ☺

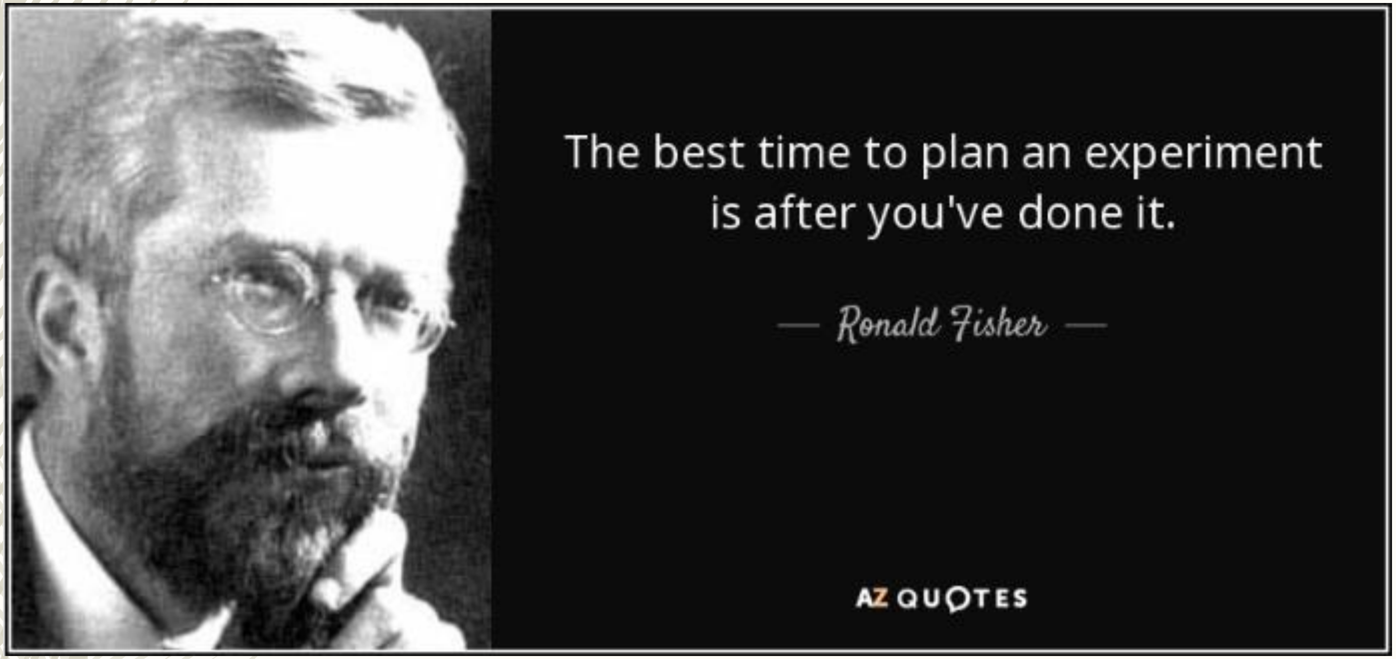
Por isso, deem o vosso papelinho com o número estimado à Susana que vai fazer um gráfico que resume a experiência.

Jelly bean challenge: resultados previstos

```
#valores estimados pelos alunos
estimativas=c(,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,)
#valor real
N = ?
#plot correspondente
hist(estimativas)
abline(v=N,col="green",lty=2)
abline(v=mean(estimativas),col="red",lty=2)
#o valor é bem estimado – se 0 estiver contido!
t.test(estimativas,mu=N)
#equivalente: o valor está contido
#no intervalo de confiança?
library(Rmisc)
CI(estimativas)
quants=quantile(estimativas,probs=c(0.025,0.975))
abline(v=quants,col="blue",lty=3)
```



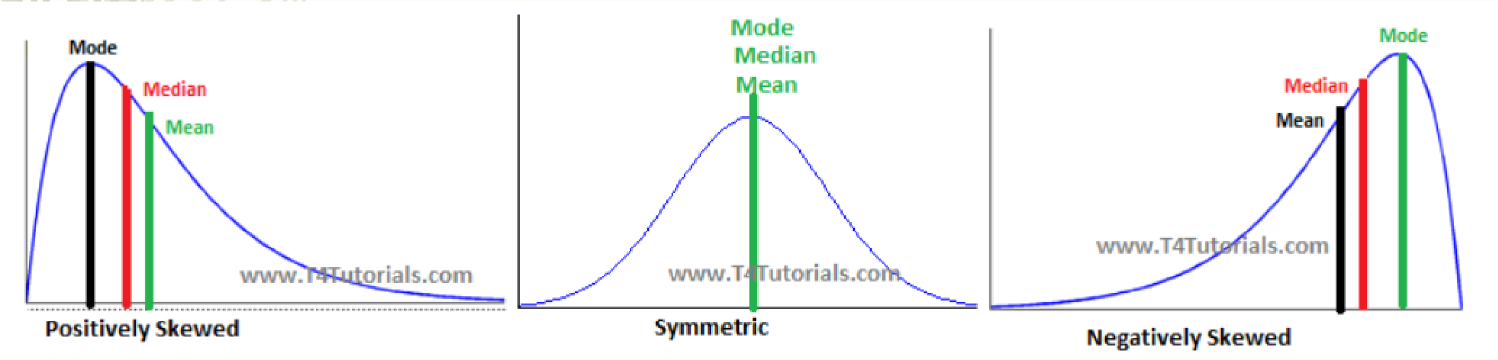
Ecologia Numérica - Aula Teórica 5 – 01-10-2018



https://www.azquotes.com/quote/1424079

Outras distribuições contínuas

- Muitas variáveis apresentam desvios à normalidade pela falta de simetria.
- Um tipo comum de desvio é o chamado *skewness* (assimetria), que é verificado quando uma das caudas da distribuição é muito mais deprimida e alongada que a outra.
- A assimetria (skewness) positiva é a mais comum (cauda alongada na parte direita da distribuição).



Famílias de distribuições assimétricas

- Existe um grande número de distribuições assimétricas, muitas das quais se enquadram na família exponencial (não confundir com a variável exponencial!)

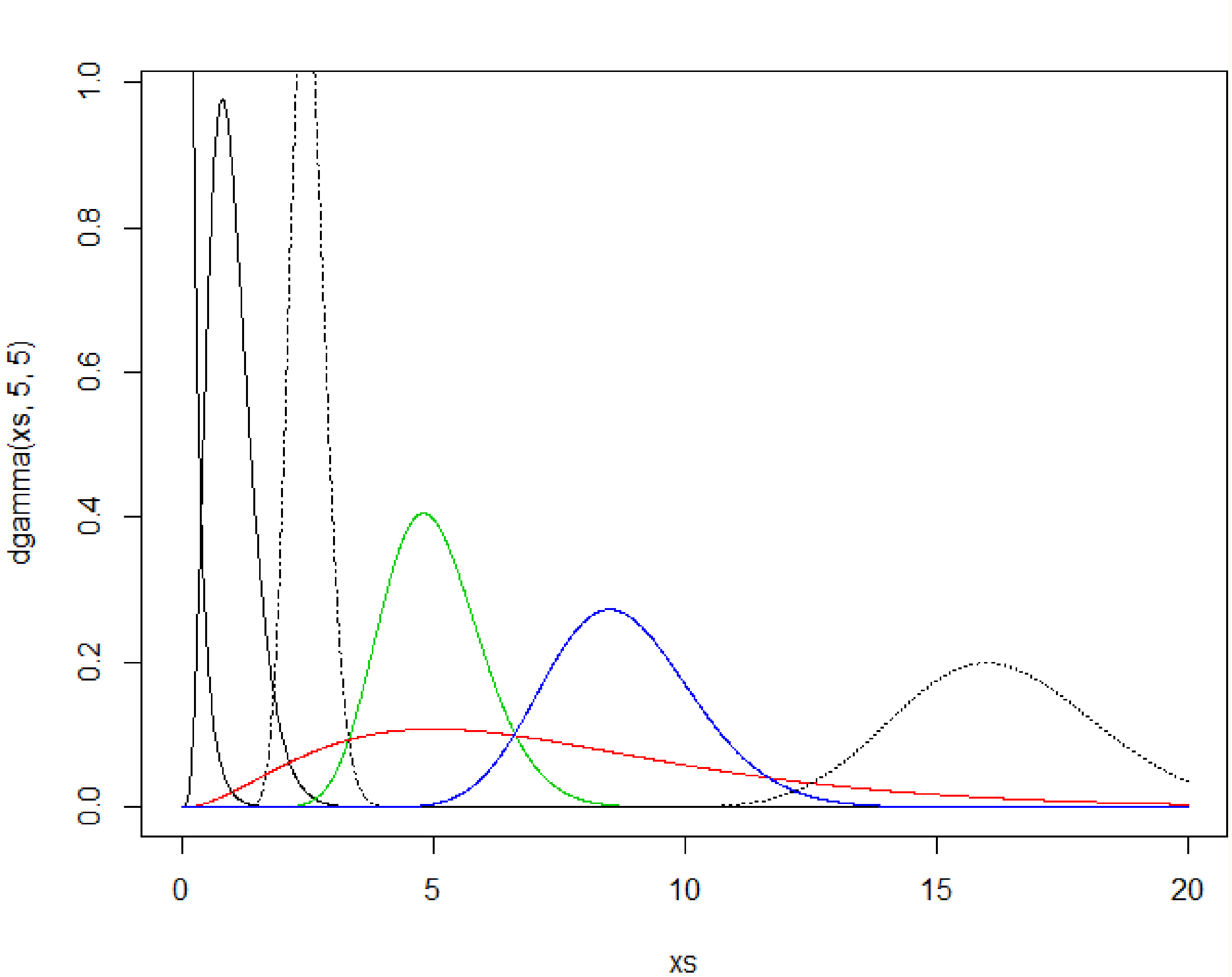
    Gamma

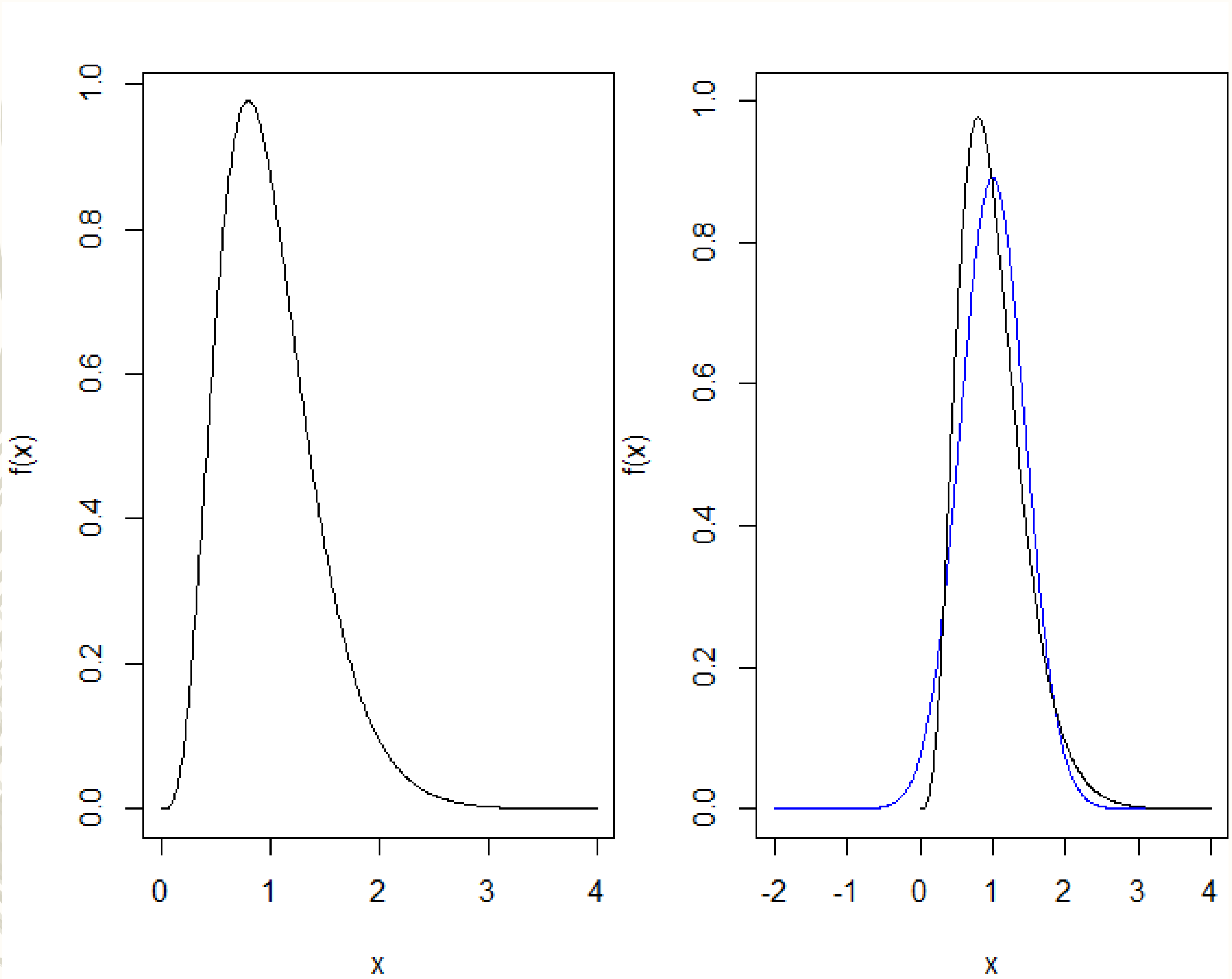
    Weibull

    Log-normal

- Estas distribuições são definidas por dois ou mais parâmetros que lhes podem conferir formas muito diferenciadas.
- São todas com suporte estritamente positivo (i.e. x>0)

Gamma - Uma distribuição com assimetria positiva



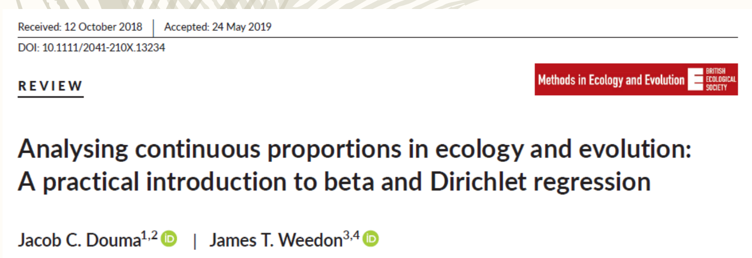


Uma Gamma(5,5) e a comparação com uma Gaussiana com a mesma média e desvio padrão

A distribuição beta

- Uma distribuição com suporte (i.e. os valores que a variável aleatória pode tomar) em (0,1)

- Uma distribuição extremamente flexível
  (e para que pode servir?)

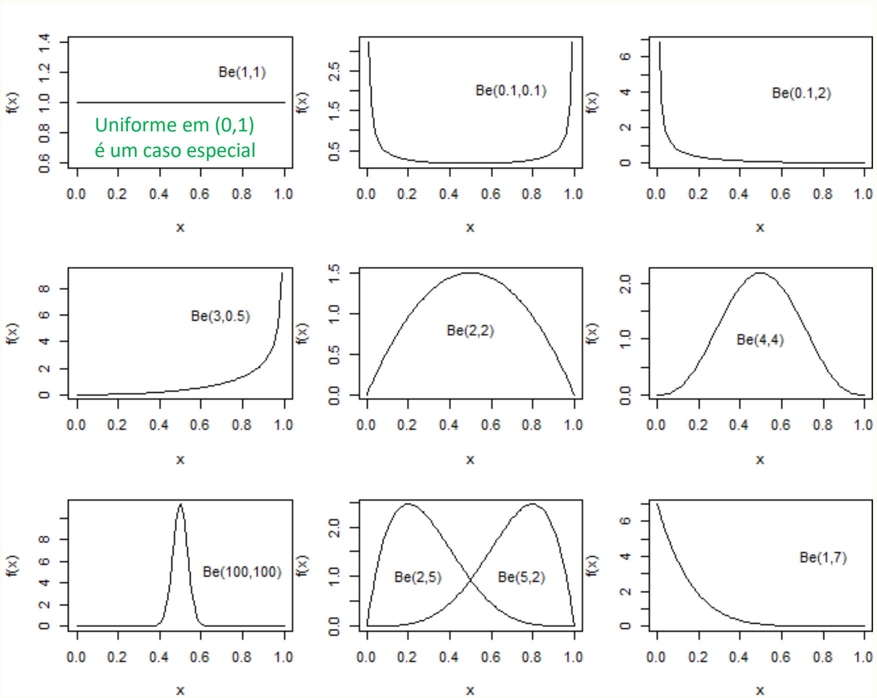

Received: 12 October 2018 | Accepted: 24 May 2019
DOI: 10.1111/2041-210X.13234

REVIEW

Methods in Ecology and Evolution

Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression

Jacob C. Douma[1,2] | James T. Weedon[3,4]



Se X for Beta então 0 < x < 1!

Probabilidades e estatísticas

A determinação de probabilidades está associada a uma população, caracterizada através duma distribuição de probababiblidades, e consiste na previsão do que poderá acontecer quando retirada uma certa amostra.

As estatísticas são determinadas a partir de amostras da população e servem para descrever os dados ou para inferir e tecer considerações sobre a população donde eram provenientes as amostras.

Distribuições no R

Distributions {stats} R Documentation

Distributions in the stats package

Description

Density, cumulative distribution function, quantile function and random variate generation for many standard probability distributions are available in the **stats** package.

Details

The functions for the density/mass function, cumulative distribution function, quantile function and random variate generation are named in the form `dxxx`, `pxxx`, `qxxx` and `rxxx` respectively.

For the beta distribution see `dbeta`.

For the binomial (including Bernoulli) distribution see `dbinom`.

For the Cauchy distribution see `dcauchy`.

For the chi-squared distribution see `dchisq`.

For the exponential distribution see `dexp`.

For the F distribution see `df`.

For the gamma distribution see `dgamma`.

For the geometric distribution see `dgeom`. (This is also a special case of the negative binomial.)

For the hypergeometric distribution see `dhyper`.

For the log-normal distribution see `dlnorm`.

For the multinomial distribution see `dmultinom`.

For the negative binomial distribution see `dnbinom`.

For the normal distribution see `dnorm`.

No R: `?Distributions`

… a lista continua, e depois há muitas outras distribuiçãos menos usuais em packages adicionais

Distribuições no R

No R: `?Distributions`

As funções:

Cada distribuição é indexada por diferentes parâmetros

Densidade:
d*****: dbinom, dpois, dnorm, dgamma, dbeta, etc…
Argumentos: quantil + parâmetros que definem a distribuição
Resultado: Qual o valor da função densidade de probabilidade no quantil q

```
dnorm(-2,mean=0,sd=1)
[1] 0.05399097
```

Função de distribuição

p*****: pbinom, ppois, pnorm, pgamma, pbeta, etc…
Argumentos: quantil + parâmetros que definem a distribuição
Resultado: Qual o valor da função de distribuição cumulativa no quantil q

```
pnorm(0,mean=0,sd=1)
[1] 0.5
```

Distribuições no R

As funções:

Função quantil

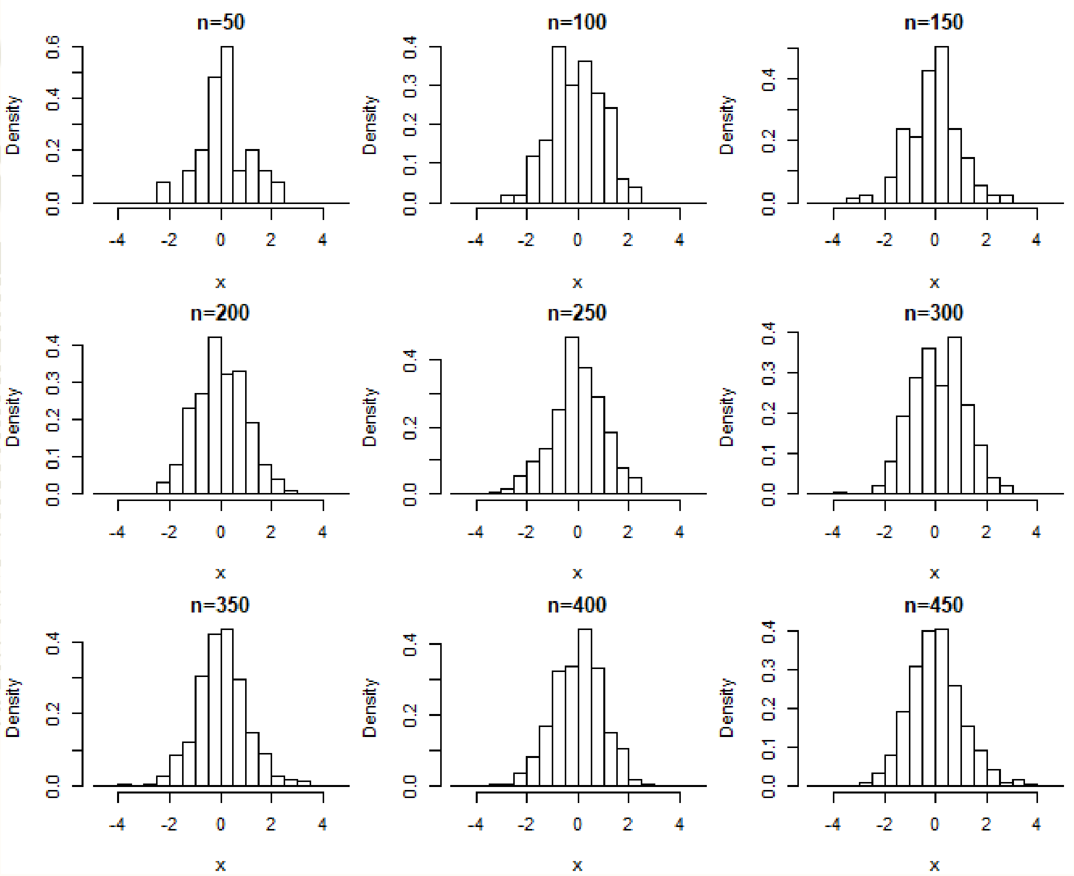
q*****: qbinom, qpois, qnorm, qgamma, qbeta, etc…
Argumentos: probabilidade + parâmetros que definem a distribuição
Resultado: Qual o quantil associado a uma determinada probabilidade

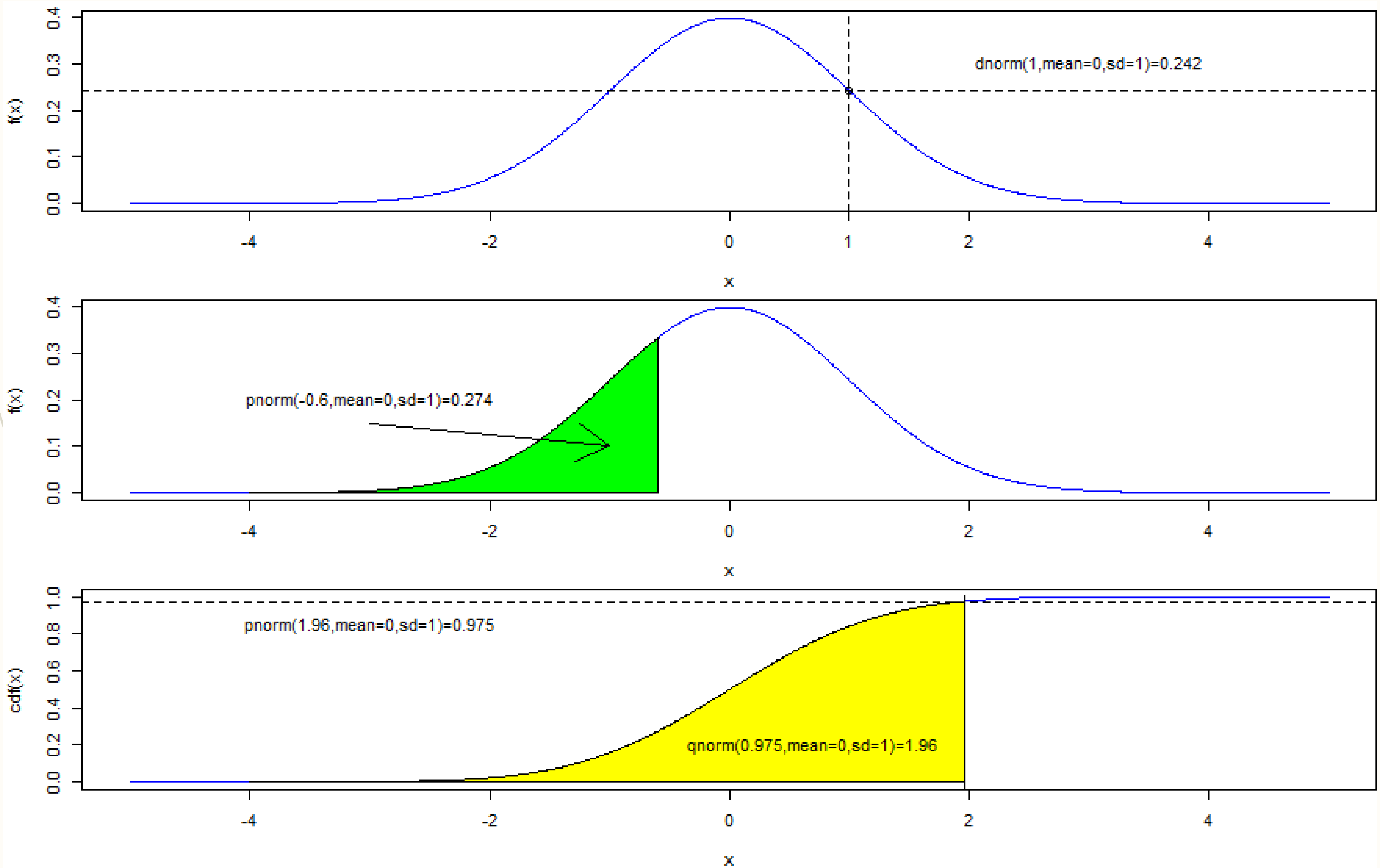
```
qnorm(0.975,mean=0,sd=1)

[1] 1.959964

qnorm(0.5,mean=0,sd=1)
[1] 0
```

Função geradora de números pseudo-aleatórios

r*****: rbinom, rpois, rnorm, rgamma, rbeta, etc…
Argumentos: número de observações + parâmetros que definem a distribuição

```{r}
set.seed(192)
par(mfrow=c(3,3),mar=c(4,4,2,0),oma=c(0,0,0,0))
for(i in 1:9){
#create histogram from n Gaussian random deviates
#with increasing sample size between 50 and 450
hist(rnorm(i*50,mean=0,sd=1),main=paste0("n=",i*50),xlab="x",
#definir que quero uma pdf
freq=FALSE,
#definir os limites das classes
breaks=seq(-5,5,by=0.5))}
```





dnorm(1,mean=0,sd=1)=0.242
pnorm(-0.6,mean=0,sd=1)=0.274
pnorm(1.96,mean=0,sd=1)=0.975
qnorm(0.975,mean=0,sd=1)=1.96
f(x)
cdf(x)
x

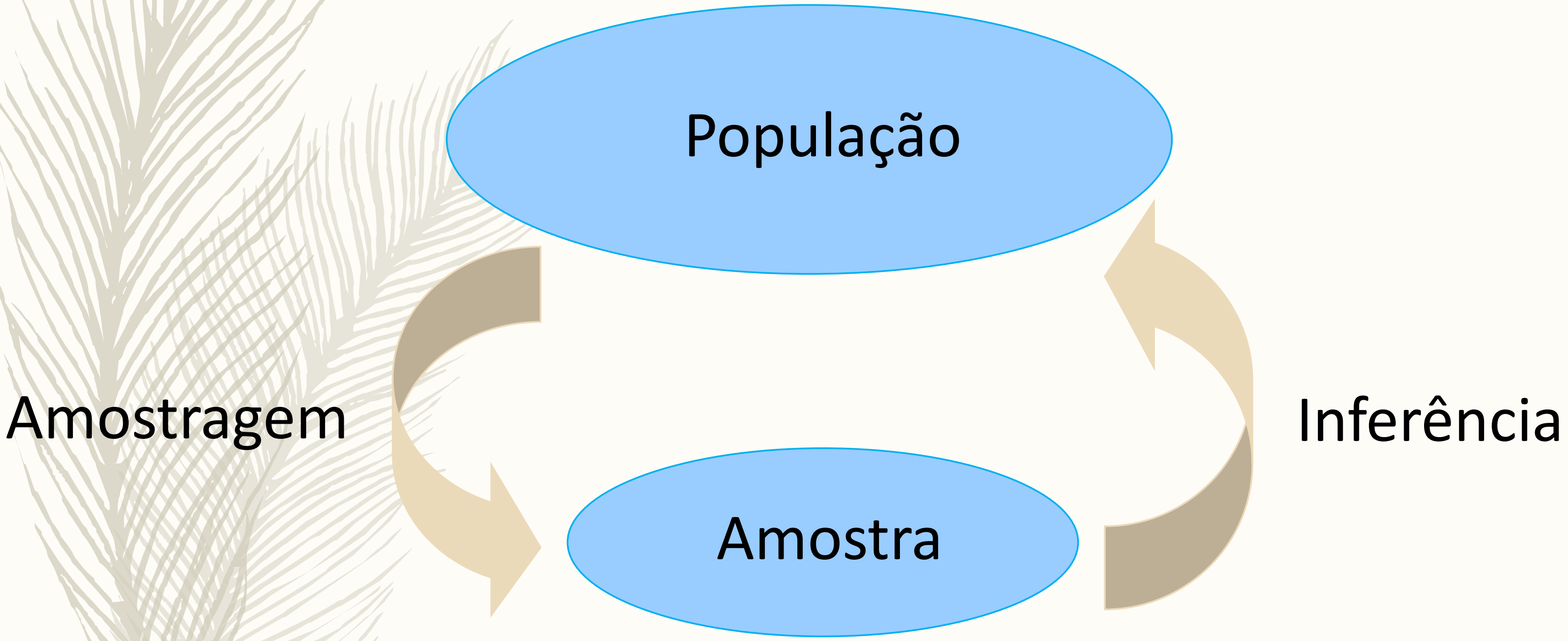
Ecologia Numérica

Amostragem e delineamento experimental

Tiago A. Marques

- Qual a necessidade de se efectuar amostragem de uma população?
- Que métodos existem?
- Que critérios são importantes serem respeitados?
- Como determinar a dimensão da amostra?
- Como desenhar uma experiência?

amostragem
delineamento experimental



Amostragem: conceitos básicos

- População (população estatística)
- Unidade de amostragem
- Método de amostragem
- Amostra (dimensão da amostra)

Métodos de amostragem:

- Amostragem aleatória simples
- Amostragem aleatória estratificada
- Amostragem sistemática
    - Outros métodos
        - (conglomerados, adaptativa, etc.)
- (estimação do tamanho de populações naturais) amostragem por distâncias, métodos de captura-recaptura e captura-recaptura espacialmente explicita

Amostragem aleatória simples

Estimativa da média

$$\bar{x} = \frac{\sum x_i}{n}$$

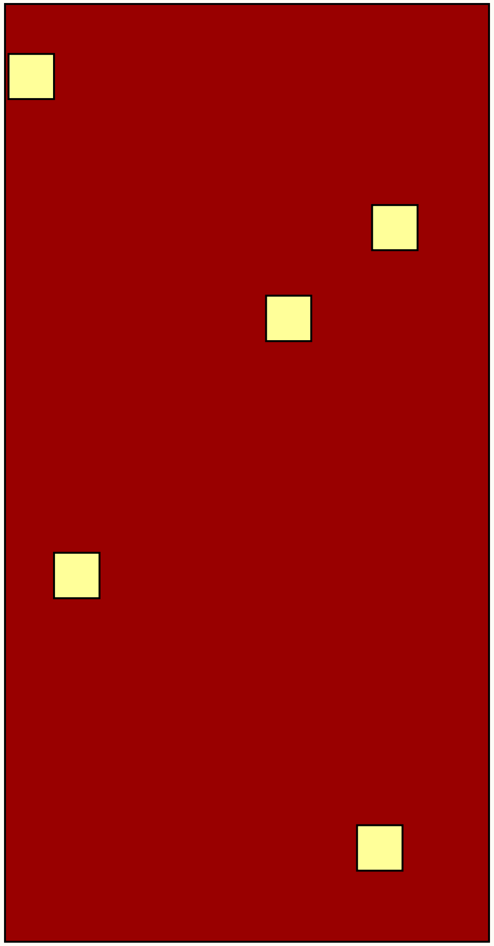
número de replicados

Erro padrão da média

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n}}\left(\sqrt{1-\frac{n}{N}}\right)$$

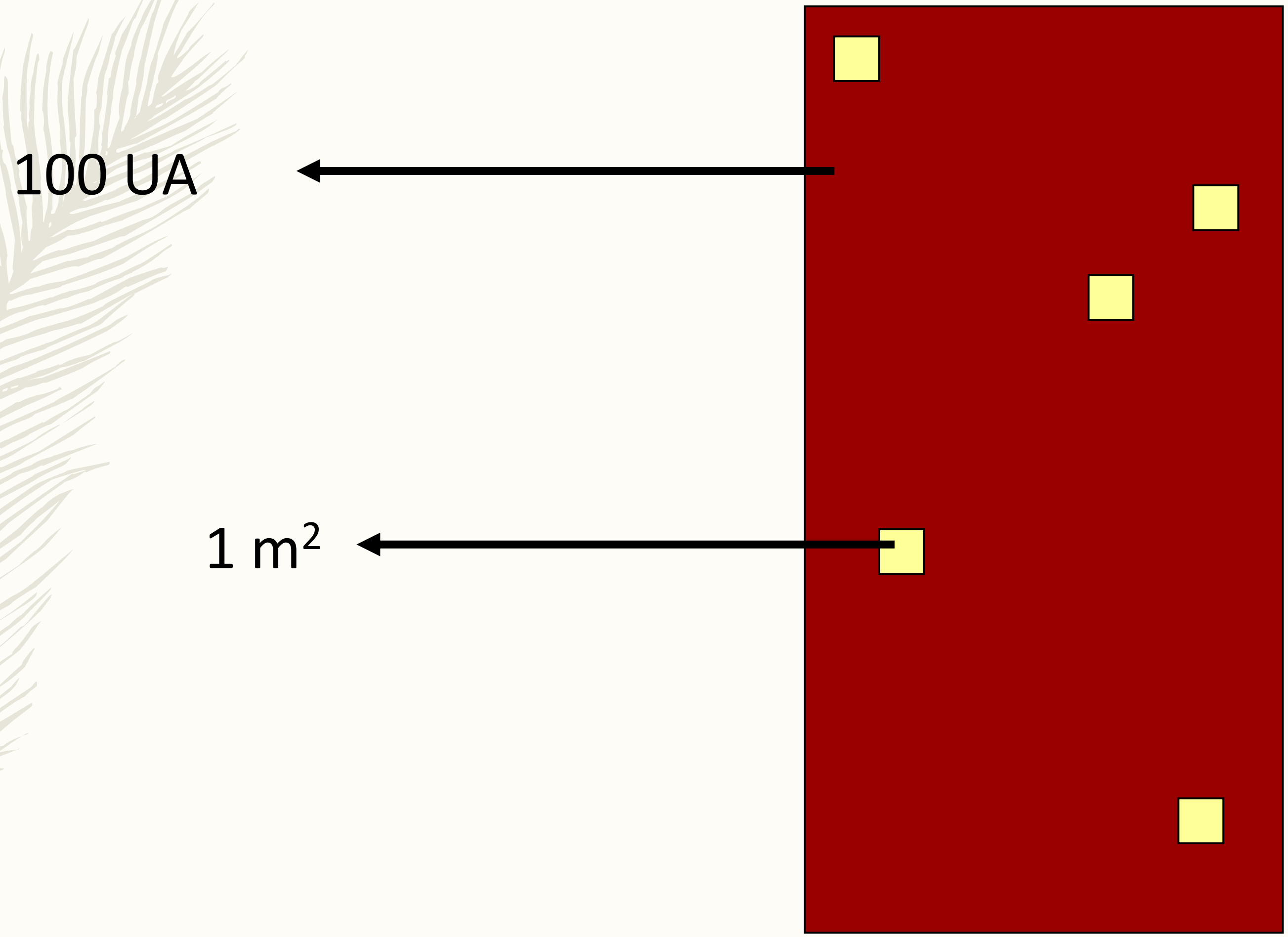
factor de correcção para populações finitas

Tamanho da população



O que acontece se n=N, i.e. se virmos tudo ?

Amostragem aleatória simples



Amostragem aleatória estratificada

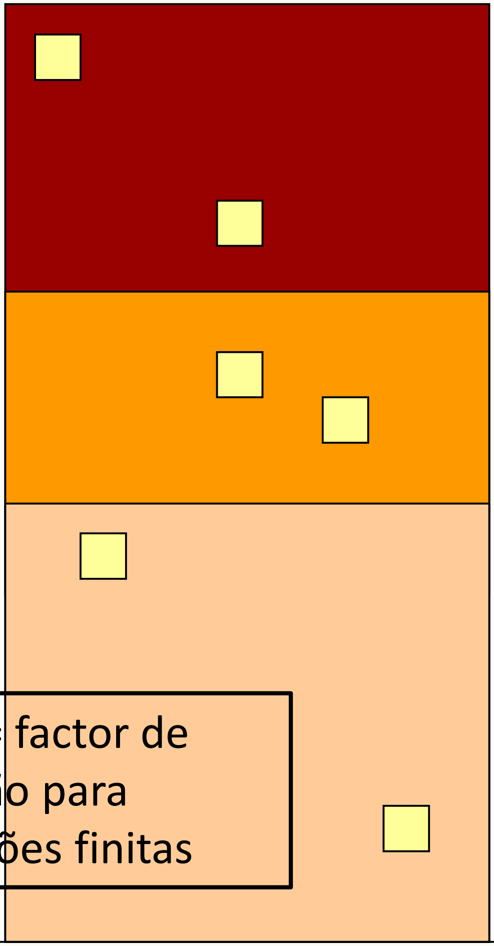
Estimativa da média

$$\bar{x}_{ST} = \frac{\sum_{i=1}^{L} N_i \bar{x}_i}{N}$$

Número de estratos (L=3 aqui ao lado)

Proporção da população no estrato i=N_i/N

onde, *N_i é a dimensão do estrato i*
e *N* a dimensão de todos os estratos



Estimativa do erro padrão

$$s_{\bar{x}} = \sqrt{\sum_{i=1}^{L}\left[\frac{w_i^2 s_i^2}{n_i}(1-f_i)\right]}$$

$1-n_i/N_i$ = factor de correcção para populações finitas

Número de replicados no estrato i

Um exemplo prático:

		4							
				11			5		
		1							
				8				4	
						6			
	3					9			
			2						
	4			1				3	
				3		5			
		0							
1									
						0			
		1							1

Estrato	Ni	ni	$\overline{xi}$	si
1	70	9	51/9	3.16
2	60	7	18/7	1.71
3	30	4	3/4	0.5
Total = N	160			

$\bar{x}_{ST} =$
```
(70*51/9+60*18/7+30*3/4)/160
[1] 3.584077
```

Estrato	fi	wi	w2*s2/n	w2*s2/n* 1-f
1	0.1	0.5	0.2787	0.25
2	0.117	0.33	0.047	0.041
3	0.133	0.17	0.0017	0.0015
sX				0.54

$S_{\bar{x}} =$
```
sqrt(sum(c(0.25,0.041,0.0015)))
[1] 0.5408327
```

Um exemplo prático:

		14							
				11			15		
		11							
				18				14	
						16			
	13					19			
			2						
	4			1				3	
				3		5			
		0							
1									
						0			
		1							1

Estrato	Ni	ni	$\bar{x}$	s
1	70	9	?/9	
2	60	7	18/7	
3	30	4	3/4	
Total = N	160			

Fazer este segundo exemplo, e comparar os resultados. Em particular, como se comparam os resultados, entre este e o primeiro exemplo de amostragem estratificada, correspondentes com os que seriam obtidos assumindo amostragem aleatória simples.